# A Trimmed Jackknife

by
David Hinkley & Hai-Li Wang
University of Minnesota
Technical Report No. 336
April, 1979

Research supported in part by NSF Grant MCS-77-00959

## Summary

The jackknifed version of a regular estimate is the average of the sample pseudo-values, which in general are not independent and often have long-tailed distributions. We examine the simple alternative of the trimmed average of the pseudo-values, which is a naive type of robust estimate. The asymptotic theory is developed and a simple example is discussed.

Key words: Jackknife; Trimmed mean; Robustness.

## 1. Introduction

The standard jackknife procedure developed by Quenouille starts wtih an estimate $T_n$ of $\theta$ based on observables $X_1,\ldots,X_n$ and, by computing pseudo-values, then constructs a modified form of $T_n$ together with a standard error estimate. The modified estimate is the simple average of pseudo-values. In a previous article (Hinkley, 1978) it was suggested that the pseudo-values might be used to construct moderately robust alternatives to $T_n$ which would be less sensitive than $T_n$ to departures from conditions under which $T_n$ is preferred. The simplest practical alternative is a trimmed mean of pseudo-values, which we call a trimmed jackknife.

In this paper we describe the asymptotic theory for the trimmed jackknife, showing that under certain regularity conditions a large-sample normal approximation exists. The variance of that approximation is consistently estimated by the Winsorized sample variance of the pseudo-values.

We suppose that the estimate $T_n$ of $\theta$ is based on ii.d. random variables $X_1,\ldots,X_n$ with common underlying distribution function $F$. Then the standard jackknife procedure is to compute pseudo-values

$$P_{n,j} = nT_n - (n-1)T_{n,-j} \quad , \quad j = 1,\ldots,n$$

where

$$T_{n,-j} = T \text{ computed from } \{ X_k : k \neq j\} \quad .$$

Then the adjusted estimate is the average $\bar{P}_n = n^{-1}\Sigma P_{n,j}$ , whose variance is estimated by (sample variance of $P_{n,j}$ ) $\div$ n . In Section 2, we develop a useful theoretical characterization of the $P_{n,j}$ when $T_n$ is a differentiable statistical function in the von Mises sense. This enables us to derive certain large-sample properties of the trimmed mean of the $P_{n,j}$ in Section 3. Some numerical illustrations are discussed in Section 4, and Section 5 contains some brief related remarks.

## 2. Characterization of Pseudo-values

In order to develop the asymptotic theory of the trimmed jackknife we need to characterize the pseudo-values in terms of von Mises functionals. First define

$$\tilde{F}_n(x) = \frac{\# X_k \leq x}{n} = n^{-1}\sum_{k=1}^{n} I(x-X_k)\ ,\ \tilde{F}_{n,-j}(x) = \frac{1}{n-1}\sum_{k\neq j} I(x-X_k)\ .$$

We shall assume that the estimate $T_n$ is of the form $t(\tilde{F}_n)$, so that the jth pseudo-value may be written

$$P_{n,j} = t(\tilde{F}_n) + (n-1)\{t(\tilde{F}_n) - t(\tilde{F}_{n,-j})\}\ . \tag{2.1}$$

(Slightly more generality is achieved by considering $T_n = t(\tilde{F}_n, z_n)$, where $z_n$ is a deterministic auxiliary variable, and this would introduce no complication into what follows.) Next suppose that for each $G$ in a compact set $\mathcal{G}$ surrounding the true distribution $F$, $t(G)$ is twice compactly differentiable. That is, for a compact set $\mathcal{H}$ surrounding $G$, the following Taylor expansion is valid

$$t(G + \overline{H-G}) = t(G) + \int t_1(G;x)\,d\overline{H-G}(x) + \iint t_2(G;x_1,x_2)\prod_{\nu=1}^{2} d\overline{H-G}(x_\nu)$$

$$+\ r(G,H-G), \tag{2.2}$$

where as $\varepsilon\to 0, r(G, \varepsilon\overline{H-G})/\varepsilon^2\to 0$ uniformly for $H\in\mathcal{H}$; see Reeds (1976). Now we can expand (2.1), using (2.2) with $G = \tilde{F}_n$ and $H = \tilde{F}_{n-1,j}$, to obtain

$$P_{n,j} = t(\tilde{F}_n) + t_1(\tilde{F}_n;X_j) - \frac{1}{2(n-1)}\,t_2(\tilde{F}_n;X_j,X_j) + (n-1)r(\tilde{F}_n,\tilde{F}_{n,-j}-\tilde{F}_n) \tag{2.3}$$

It is important to note here that $t_1$ and $t_2$ have been standardized so that $\int t_1(F;x)dF(x) = 0$ , $\int t_2(F;x,y)dF(x) \equiv 0$ , and therefore the first derivative of $t_1(F;x)$ at $(F,y)$ is

$$t_{1,1}(F;x;y) = t_2(F;x,y) - t_1(F;y) \quad . \tag{2.4}$$

As regards the final term in (2.3) notice that

$$\max_{1 \leq j \leq n} \sup_x |\tilde{F}_n(x) - \tilde{F}_{n,-j}(x)| = \frac{1}{n} \quad ,$$

which implies that the remainder term in (2.3) satisfies

$$\max_{1 \leq j \leq n} (n-1) r(\tilde{F}_n, \tilde{F}_{n,-j} - \tilde{F}_n) = 0_p(n^{-1}) \quad . \tag{2.5}$$

Since the next-to-last term in (2.3) is $0_p(n^{-1})$ we conclude that

$$P_{n,j} = t(\tilde{F}_n) + t_1(\tilde{F}_n;X_j) + 0_p(n^{-1}) \quad . \tag{2.6}$$

But we can go further if for $G \varepsilon \mathcal{J}$ both $t_2(G;x,x)$ and $t_1(G;x)$ are regular functions of $x$ , since then for any $j,k$

$$\frac{\frac{1}{n-1} t_2(\tilde{F}_n;X_j,X_j) - \frac{1}{n-1} t_2(\tilde{F}_n;X_k,X_k)}{t_1(\tilde{F}_n;X_j) - t_1(\tilde{F}_n;X_k)} = 0_p(n^{-1}) \quad . \tag{2.7}$$

This, together with (2.5), will imply that (2.6) is true for <u>ordered</u> values of $P_{n,j}$ , i.e. with obvious notation

$$P_{(n,j)} = t(\tilde{F}_n) + \{ t_1(\tilde{F}_n;X)\}_{(n,j)} + 0_p(n^{-1}) \quad . \tag{2.8}$$

Essentially (2.5) guarantees that the final term in (2.3) has no effect on the ordering because differences between $P_{(n,j)}$ are of order $n^{-1}$ . Then (2.7) guarantees that the $t_2$ term in (2.3) ultimately does not disturb the relative orderings of $P_{n,j}$ and $t(\tilde{F}_n) + t_1(\tilde{F}_n;X_j)$ .

The representations (2.6) and (2.8) are valuable in studying the asymptotic properties of jackknife statistics, although the additional $t_2$ term in (2.3) is required to distinguish between the jackknife and the infinitesimal jackknife.

## 3. The Trimmed Jackknife and Its Asymptotic Distribution.

For several common estimation problems the jackknife pseudo-values $P_{n,j}$ are approximately symmetric in distribution because $t_1(F;X)$ has a symmetric distribution. Particular instances include correlation estimation and estimation of functions of regression parameters for models with symmetric errors. In such cases it is not unnatural to consider symmetric estimates other than $\bar{P}_n$ which de-emphasise large deviations $P_{n,j} - T_n$. The simplest alternative would seem to be a trimmed mean

$$\frac{1}{n-2r} \sum_{j=r+1}^{n-r} P_{(n,j)}$$

for some $r \geq 1$, which should be robust to moderate deviations from the sampling situation where $\bar{T}_n$ is the estimate of choice. In this section we show that this trimmed jackknife has an asymptotic normal distribution, with correct mean $\theta$ if $t_1(F;X)$ is symmetrically distributed, and with a variance that can be consistently estimated by the Winsorized sample variance of the $P_{n,j}$. We shall assume that the number of pseudo-values trimmed at each end of their ordered list is $r = [\alpha n]$, the integer part of $\alpha$ times sample size.

Armed with (2.8) we are now in a position to characterize the trimmed jackknife

$$T_n^{(\alpha)} = \frac{1}{n(1-2\alpha)} \sum_{j=r_\alpha+1}^{n-r_\alpha} P_{(n,j)} \quad , \quad r_\alpha = [n\alpha]$$

in functional form, following which a von Mises delta method will give the limiting distribution. We write

$$T_n^{(\alpha)} = t^{(\alpha)}(\tilde{F}_n) + 0_p(n^{-1})$$

with

$$t^{(\alpha)}(\tilde{F}_n) = t(\tilde{F}_n) + \frac{1}{n(1-2\alpha)} \sum_{j=r_\alpha+1}^{n-r_\alpha} \{ t_1(\tilde{F}_n;X)\}_{(n,j)} \quad . \tag{3.1}$$

It is convenient to define

$$\tilde{K}_n(y) = \frac{\#\{ T_n+t_1(\tilde{F}_n;X_j) \leq y\}}{n} = \int I\{ y-t(\tilde{F}_n)-t_1(\tilde{F}_n;x)\}d\tilde{F}_n(x) \quad ,$$

which by convergence of $\tilde{F}_n$ to $F$ converges to

$$K(y) = \mathrm{pr}\{ t(F) + t_1(F;X) \leq y\} \quad .$$

Then the function (3.1) may be written

$$t^{(\alpha)}(F) = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} K^{-1}(u)du \quad , \tag{3.2}$$

where of course $K$ and $K^{-1}$ are functions of $F$ .

If we denote the (centered) derivative of $t^{(\alpha)}(F)$ by $t_1^{(\alpha)}(F;x)$ , then the von Mises delta method gives

$$T_n^{(\alpha)} = t^{(\alpha)}(\tilde{F}_n) + 0_p(n^{-1}) = t^{(\alpha)}(F) + n^{-1} \sum_{j=1}^{n} t_1^{(\alpha)}(F;X_j) + 0_p(n^{-1}) \ ,$$

so that as $n\to\infty$

$$\sqrt{n}\ (T_n^{(\alpha)} - t^{(\alpha)}(F)) \sim N(0,\sigma^2)$$

with

$$\sigma^2 = \mathrm{Var}\{ t_1^{(\alpha)}(F;X)\} \quad . \tag{3.3}$$

Evidently from (3.2) $t^{(\alpha)}(F) = \theta$ if $t_1(F;X)$ is symmetrically distributed. It now remains only to compute $t_1^{(\alpha)}(F;x)$ and thence $\sigma^2$ or an estimate of $\sigma^2$ .

Let the first derivative of $K(s)$ and $J(u) = K^{-1}(u)$ w.r.t. $F$ at $(F,x)$ be denoted by $K_1(s;x)$ and $J_1(u;x)$ respectively, and let the density of $K(s)$ be $k(s)$ . Then it is easy to show that

$$J_1(u,x) = -K_1\{ K^{-1}(u);x\} \ / \ k\{ K^{-1}(u)\} \ .$$

Next, using (2.4) together with the expansion (2.2) for $K(s)$ , we find that

$$K_1(s;x) = I\{ s-t(f)-t_1(F;x)\} - K(s)$$
$$- k(s)E\{ t_2(F;x,Y)\,|\,t_1(F;Y) = s\} \quad .$$

Therefore

$$J_1(K(s);x) = \frac{K(s)-I\{ s-t(F)-t_1(F;x)\}}{k(s)} + E\{ t_2(F;x,Y)\,|\,t_1(F;Y) = s\} \tag{3.4}$$

For the linear function $t^{(\alpha)}(F)$ defined in (3.2) we use (3.4) to obtain

$$t_1^{(\alpha)}(F;x) = \frac{1}{1-2\alpha}\int_{\alpha}^{1-2\alpha} J_1(u;x)du$$

$$= \frac{1}{1-2\alpha}\int_{K^{-1}(\alpha)}^{K^{-1}(1-\alpha)} [K(s)-\dot{I}\{s-t(F)-t_1(F;x)\}]ds$$

$$+ \frac{1}{1-2\alpha}\int_{K^{-1}(\alpha)}^{K^{-1}(1-\alpha)} \dot{E}\{t_2(F;x,Y)\,|\,t_1(F;Y)=s\}k(s)ds \ . \qquad (3.5)$$

If we now assume that $K(.)$ is symmetric, (3.5) simplifies to

$$t_1^{(\alpha)}(F;x) = [t_1(F;x)]_{\alpha}^{1-\alpha} + \dot{E}\{t_2(F;x,Y)\,|\,K^{-1}(\alpha) \leq t_1(F;Y) \leq K^{-1}(1-\alpha)\} \ , \qquad (3.6)$$

where

$$(1-2\alpha)[t_1(F;x)]_{\alpha}^{1-\alpha} = \begin{cases} K^{-1}(1-\alpha) & K^{-1}(1-\alpha) \leq t_1(F;x) \ , \\ t_1(F;x) & K^{-1}(\alpha) \leq t_1(F;x) < K^{-1}(1-\alpha) \ , \\ K^{-1}(\alpha) & t_1(F;x) \leq K^{-1}(\alpha) \ . \end{cases}$$

Notice the similarity between (3.6) and the influence function for the ordinary trimmed mean, the difference here being the occurrence of the "non-linearity" term involving $t_2$ . The magnitude of this additional term clearly depends on $F$ .

Given the above form for $t_1^{(\alpha)}(F;x)$ we can in principle compute its variance $\sigma^2$. Pratical interest requires rather an estimate of $\sigma^2$ , for which the obvious candidate is

$$\tilde{\sigma}_n^2 = n^{-1}\sum_{j=1}^{n}\{t_1^{(\alpha)}(\tilde{F}_n;X_j)\}^2 \ . \qquad (3.7)$$

Calculation of $t_1^{(\alpha)}$ , and in particular of $t_2$ , can be avoided by noting that under our differentiability assumptions $\sigma^2$ is also consistently estimated by the jackknife variance estimate

$$S_n^{(Q)2} = n^{-1} \sum_{j=1}^{n} (Q_{n,j} - \bar{Q}_n)^2 \tag{3.8}$$

where $Q_{n,j}$ is the jth pseudo-value of $t^{(\alpha)}(\tilde{F}_n)$ and $\bar{Q}_n = n^{-1}\Sigma Q_{n,j}$ . This is essentially guaranteed by relating $Q_{n,j}$ to $t_1^{(\alpha)}(\tilde{F}_n;X_j)$ as in (2.8). If we now ignore the $o_p(1)$ term arising from the difference $\{t_1(\tilde{F}_n;X)\}_{(n,k)} - \{t_1(\tilde{F}_{n,-j};X)\}_{(n,k)}$ and write $R_j = \text{rank}\,\{t_1(F_n;X_j)\}$ , then we find

$$(1-2\alpha)Q_{n,j} = \begin{cases} t(\tilde{F}_n) + \{t_1(\tilde{F}_n;X)\}_{(n,r_\alpha+1)} & (R_j \leqslant r_\alpha) \\ t(\tilde{F}_n) + t_1(\tilde{F}_n;X_j) & (r_\alpha+1 \leqslant R_j \leqslant n-r_\alpha) \\ t(\tilde{F}_n) + \{t_1(\tilde{F}_n;X)\}_{(n,n-r_\alpha)} & (n-r_\alpha+1 \leqslant R_j) \end{cases} .$$

But by (2.6) and (2.8) this gives

$$(1-2\alpha)Q_{n,j} = \begin{cases} P_{(n,r_\alpha+1)} + o_p(1) & (\text{rank}(P_{n,j}) \leqslant r_\alpha) \\ P_{n,j} + o_p(1) & (r_\alpha+1 \leqslant \text{rank}(P_{n,j}) \leqslant n-r_\alpha) \\ P_{(n,n-r_\alpha)} + o_p(1) & (n-r_\alpha+1 \leqslant \text{rank}\,(P_{n,j})) \end{cases} .$$

Therefore the estimate (3.8) is approximately equal to the usual Winsorized sample variance of the $P_{n,j}$ , namely

$$S_n^{(P)2} = \frac{1}{(1-2\alpha)^2} \{\frac{1}{n} \sum_{j=r_\alpha+1}^{n-r_\alpha} (P_{(n,j)}-T_n^{(\alpha)})^2 + \alpha(P_{(n,r_\alpha+1)}-T_n^{(\alpha)})^2 + \alpha\,(P_{(n,n-r_\alpha)} - T_n^{(\alpha)})^2\} . \tag{3.9}$$

This is the most convenient estimate of $\sigma^2$ , but not necessarily the most reliable, since several approximations are involved.

Taking for granted the fact that $S_n^{(P)2} = \sigma^2 + o_p(1)$ , our summary result is that for symmetric differentiable $t_1(F;X)$ , as $n\to\infty$

$$\sqrt{n}(T_n^{(\alpha)} - \theta)/S_n \sim N(0,1)$$

for each of $S_n = \tilde{\sigma}_n, S_n^{(Q)}, S_n^{(P)}$ defined respectively in (3.7), (3.8) and (3.9).

A final important point: note from (3.6) that the influence function of $T_n^{(\alpha)}$ is not bounded, owing to the term in $t_2$ . In fact (3.6) may be seen to match the influence function given by Hinkley (1978) for the appropriate Huber Type I M-estimate based on the $P_{n,j}$ . However, $T_n^{(\alpha)}$ does severely restrict the influence for points $x$ having large values of $t_1$ .

The next section describes some numerical results relating to an application of the trimmed jackknife.

## 4. Numerical Illustration

We use the correlation example from Hinkley (1978) to illustrate the performance of the trimmed jackknife. For bivariate data $X=(Y,Z)$ , $T_n$ is the Fisher-transformation of the sample correlation, $T_n = \tanh^{-1} R_n$, estimating $\theta = \tanh^{-1}\rho$ . In this case, if we write $\tilde{Y}$ and $\tilde{Z}$ for standardized variables, then

$$t_1(F;x) = \{ \tilde{y}\tilde{z} - \tfrac{1}{2}\rho\ (\tilde{y}^2+\tilde{z}^2)\}/(1-\rho^2) \quad .$$

This has a symmetric distribution if the distribution of X is elliptically symmetric, e.g. a mixture of bivariate normals.

First consider the three samples of size $n = 20$ scatter-plotted in Figure 1. These samples, used in Section 2 of Hinkley (1978), have nineteen pseudo-normal pairs in common, and differ in respect of $x_{20}$ whose values are $(0,0)$, $(0.5,-0.5)$ and $(1.0,-1,0)$. Table 1b gives values of trimmed means $T_n^{(\alpha)}$ and Winsonized sample variances $S_n^{(P)2}$ for $\alpha = 0, 0.05$ and $0.10$ in each sample, together with approximate 95% confidence limits for $\rho$ based on transformation of

$$T_n^{(\alpha)} \pm 2\, S_n^{(P)}/\sqrt{n} \quad .$$

Apparently the trimmed-mean analyses are reasonably resistant to gross outliers, in the sense that they agree well with the usual analysis of the "normal" first sample $(x_{20} = 0)$.

A few comments are in order here. First, the trimmed mean analysis pre-supposes that we wish to estimate the correlation of the central part of the data with others suppressed. It would be usual then to indicate the outliers, which here suggest themselves by inspection of the values of $P_{n,j} - T_n$ or, better, $P_{n,j} - T_n^{(\alpha)}$ . Second, the construction of $T_n^{(\alpha)}$ requires no explicit reference to other parameters, in this case means and variances. Many methods would require simultaneous robust estimation of these other parameters, possibly with a resultant gain in robustness.

## 5. Further Remarks

The asymptotic behaviour of the trimmed jackknife described in Section 3 is a mild generalization of what is known about trimmed means. Comparison of influence function (3.6) with that for the Huber Type 1 M-estimate (Section 4 of Hinkley, 1978) shows that Jaeckel's (1971) duality result extends to this more general set-up.

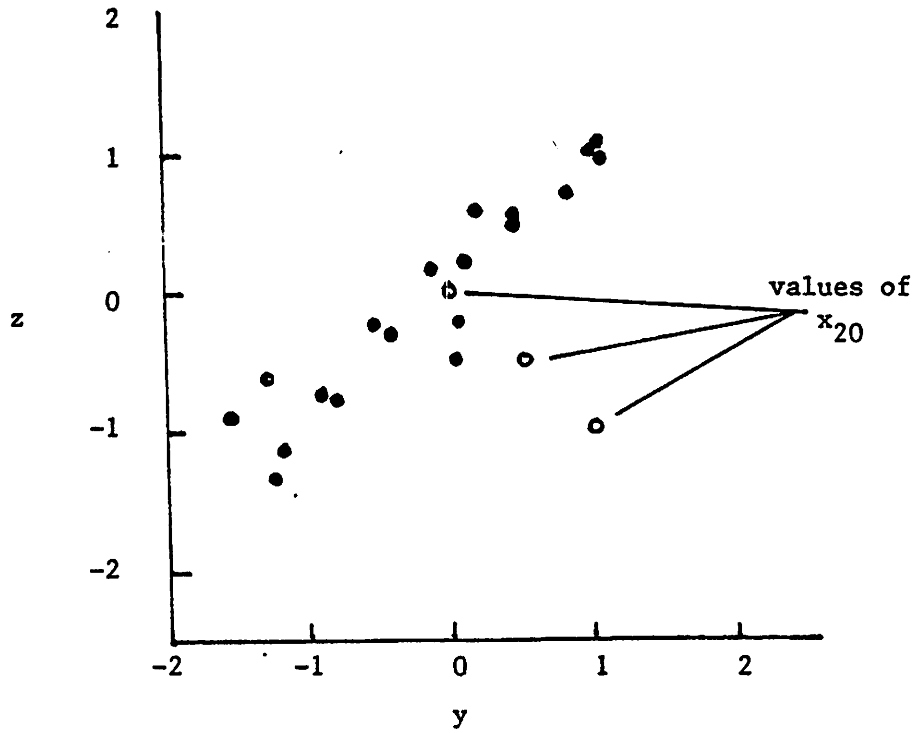


FIG.1 Three artificial bivariate samples of size n=20. Nominal distribution of $x_1,\ldots,x_{19}$ is bivariate normal with correlation 0.95 and N(0,1) margins.

Table 1a. Pseudo-values $P_{n,j}$ of $T_n = \tanh^{-1}(R_n)$ for three samples graphed in Figure 1

| j | $x_{20}$ = | (0,0) | (0.5,-0.5) | (1.0,-1.0) |
|---|---|---|---|---|
| 1 | | 0.62 | 0.67 | 0.67 |
| 2 | | -1.39 | -0.60 | 0.02 |
| 3 | | -0.83 | -0.53 | -0.20 |
| 4 | | -0.52 | 0.11 | 0.61 |
| 5 | | -0.15 | -0.12 | -0.06 |
| 6 | | 1.17 | 1.17 | 1.16 |
| 7 | | 0.05 | 0.03 | 0.03 |
| 8 | | 0.88 | 1.01 | 1.06 |
| 9 | | 0.44 | 1.01 | 1.26 |
| 10 | | 1.15 | 1.15 | 1.15 |
| 11 | | 0.06 | 0.05 | 0.04 |
| 12 | | 0.47 | 0.48 | 0.47 |
| 13 | | 0.34 | 0.42 | 0.44 |
| 14 | | 0.29 | 0.27 | 0.27 |
| 15 | | -1.03 | -0.40 | -0.12 |
| 16 | | 1.22 | 1.20 | 1.20 |
| 17 | | -3.87 | -1.58 | -0.55 |
| 18 | | 0.28 | 0.26 | 0.26 |
| 19 | | -0.33 | -0.27 | -0.17 |
| 20 | | -0.04 | -5.88 | -13.54 |

Table 1b. Trimmed jackknife $T_n^{(\alpha)}$, Winsonized sample variance $S_n^{(P)2}$ and approximate 95% confidence limits for $\rho$ based on pseudo-values in Table 1a for $\alpha$=0, 0.05, 0.10

| | | $x_{20} = (0,0)$ | $x_{20} = (0.5,-0.5)$ | $x_{20} = (1.0,-1.0)$ |
|---|---|---|---|---|
| $\alpha = 0$ | $T_n^{(\alpha)} =$ | 1.70 | 1.37 | 0.75 |
| | $S_n^{(P)2} =$ | 1.25 | 2.25 | 9.51 |
| | 95% limits = for $\rho$ | 0.83,0.98 | 0.60,0.97 | -0.56,0.97 |
| $\alpha = 0.05$ | | 1.84 | 1.62 | 1.40 |
| | | 0.73 | 0.78 | 0.39 |
| | | 0.90,0.98 | 0.84,0.97 | 0.81, 0.93 |
| $\alpha = 0.10$ | | 1.87 | 1.67 | 1.40 |
| | | 0.77 | 0.59 | 0.40 |
| | | 0.90, 0.98 | 0.87, 0.97 | 0.81, 0.93 |

Is the trimmed jackknife useful? In the sense that it is a general method for de-sensitizing estimates, the answer is yes -- with the qualification that consistent results require a certain symmetry of distribution, as described in Section 3. The idea of the trimmed jackknife may also be applied to non-homogeneous situations, as mentioned by Hinkley (1977), but care is needed. A similar all-purpose approach to "robustifying" estimates $T_n$ is to define estimates directly in terms of the influence function of $T_n$. For example, write $a(\theta,w;x) = t_1(f;x)$ to emphasize the parametization of $F$, with $w$ a nuisance parameter. Then one may define an M-estimate $T_n^{(\psi)}$ as a solution to $\Sigma\psi\{a(t,\hat{w};X_j)\} = 0$, for suitable choice of $\psi(.)$ and estimate $\hat{w}$. This proposal is studied in detail by Wang & Hinkley (1979).

In any given problem, it may be possible to construct more reliable estimates. This is the case for correlation estimation. The more realistic answer to our questions is: yes, as a simple but effective complement to the jackknife procedure.

The research described in this paper is an outgrowth of work in the second author's Ph.D. thesis at the University of Minnesota. Partial support by the National Science Foundation is gratefully acknowledged.

## References


Hinkley, D.V. (1977). Jackknifing in unbalanced situations. Technometrics, 19, 285-292.

Hinkley, D.V. (1978). Improving the jackknife with special reference to correlation estimation. Biometrika, 65, 13-21.

Jaeckel, L.A. (1971). Robust estimates of location: symmetry and asymmetric contamination. Ann. Math. Stat. 42, 1020-1034.

Reeds, J.A. (1976). On the definition of von Mises functionals. Harvard University Ph.D. thesis.

Wang, H.L. & Hinkley, D.V. (1979). Robust estimation using influence function as error. University of Minnesota School of Statistics Technical Report No. 337.